Cosmic Emulation: The Universe as a *Black Box*

Earl Lawrence, CCS-6; Katrin Heitmann, ISR-1; David Higdon, CCS-6;
Christian Wagner, ISR-1; Martin White, University of California, Berkeley;
Salman Habib, T-2; Brian Williams, CCS-6

Cosmology is perhaps the grandest of inverse problems: given knowledge of observational data, we wish to infer the fundamental laws governing the dynamics of the Universe and the properties of its constituents. The difficulty, of course, is that the Universe is very complicated and very large, so that the cosmic inverse problem is by no means easy to solve.

However there is one regime where one can do well, and that is studying the Universe on very large scales. For example, the primary temperature anisotropies in the cosmic microwave background (CMB) are small and can be measured accurately on angular scales large enough so that a linear treatment of cosmological fluctuations suffices to describe them. As a result, the associated inverse problem – determination of a handful of cosmological parameters – can be solved in a relatively straightforward manner.

The story becomes much more complicated when tracers of the distribution of matter are involved and when one goes to small length scales where the physics becomes nonlinear and complex. There are two reasons why this ostensibly more complex regime is nevertheless of very significant interest. First, investigations of the structure of the Universe on large scales suffer from the fact that the associated statistical error bars are large due to unavoidable finite sampling limitations, the so-called *cosmic variance* problem. Fluctuations on small scales are much better sampled, with negligible cosmic variance. Second, it is very important to have multiple probes that can be crosschecked with each other in order to avoid systematic biases, and to include information missing from a certain cosmological probe, or class of probes, due to parametric degeneracies.

Solving the cosmic inverse problem has gained significant urgency with the discovery that the expansion of the Universe is actually accelerating, rather than slowing down. Either a very strange new component is driving this acceleration, or general relativity – our cherished theory of gravitation – may need to be modified. It is widely thought that cosmological measurements (and their interpretations) at the percent level of accuracy are required to begin to answer some of these questions. Because of the severe computational difficulties, as well as scientific uncertainties, large-scale numerical simulations and phenomenological approaches are the only tools available to address the inverse problem.

The classic approach to a statistical inverse problem is via Bayesian inference, where the posterior distribution is sampled by Markov Chain Monte Carlo (MCMC) methods. This approach involves solving the forward model tens to hundreds of thousands of times. If the forward model is a complicated numerical code that takes days to run, the MCMC approach becomes impractical. To overcome this problem, we have introduced a framework that we call *cosmic calibration* [1,2]. An essential aspect of our approach is the use of interpolation in high dimensions, via principal components and Gaussian process modeling, to generate forward model predictions while having run only a small number of base simulations (of the order of tens to hundreds). We refer to the interpolation system as a forward model *emulator*.

A high-accuracy demonstration of this approach is presented in a sequence of three papers targeting precision emulation of the mass fluctuation power spectrum down to scales small enough to be relevant to current weak gravitational lensing surveys mapping the distribution of matter. The first paper [3] shows that the underlying simulations are accurate to the 1% level, the second [4] shows that the emulation scheme is also as accurate (by using an approximate stand-in model for the actual simulations), and the third paper [5] presents the full emulator. The remarkable aspect of having an operational emulator is that it replaces simulations of the Universe by an effectively infinitely fast *black box* oracle. Using the oracle can reduce the MCMC analysis times from years to minutes.

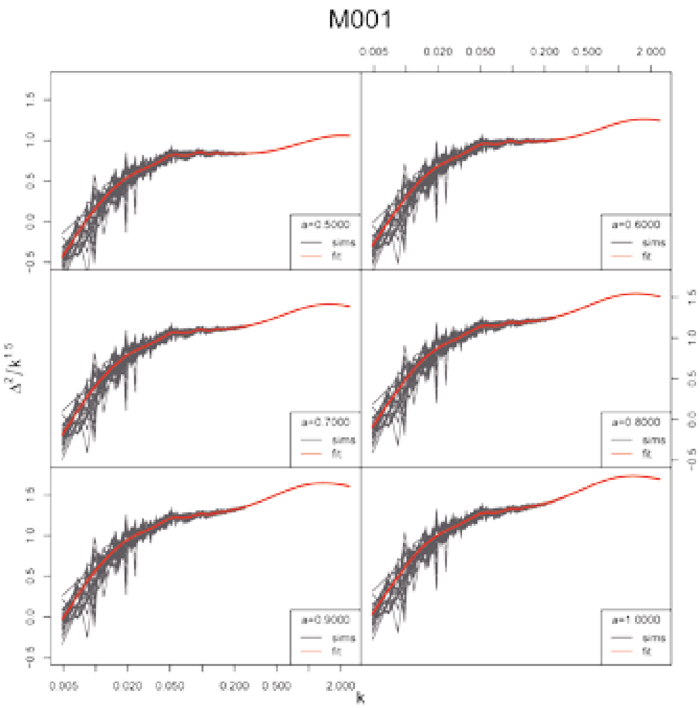

Fig. 1. Simulated dimensionless power spectra for the mass fluctuations in one of the 37 simulated cosmological models. The large fluctuations at small values of the wave number are due to cosmic variance, perturbation theory is used to fit the k< 0.03 Mpc^{-1} regime. At larger values of k, process convolution is used to smooth the simulation data—note that subtle features such as the small baryon acoustic oscillation wiggles are nicely fit by this technique.

The construction of the cosmic *black box* was accomplished by running a suite of multiresolution simulations on LANL's Coyote cluster. Almost 40 cosmologies were run with roughly 20 simulations per cosmology, for a total of almost 1000 large-scale simulations and an associated dataset size of 60 terabytes (TB), the largest simulation suite of its kind. The parameters for the cosmological models were sampled using Latin Hypercube sampling (for details, see [4]) over a prior range consistent with current observational constraints. Figure 1 shows the dimensionless power spectrum for one of the simulated cosmologies at six values of the expansion factor. The smooth fit is obtained by combining perturbation theory at large scales and adaptively filtering the noisy simulation results at smaller scales using process convolution. Figure 2 demonstrates the accuracy of the emulation technique. Although only 37 sampling points were used over a five-dimensional space, the emulation accuracy for predicting a function over more than two decades of dynamic range is (mostly) better than a percent. We have made the emulator publicly available so that it can be used by the entire community.

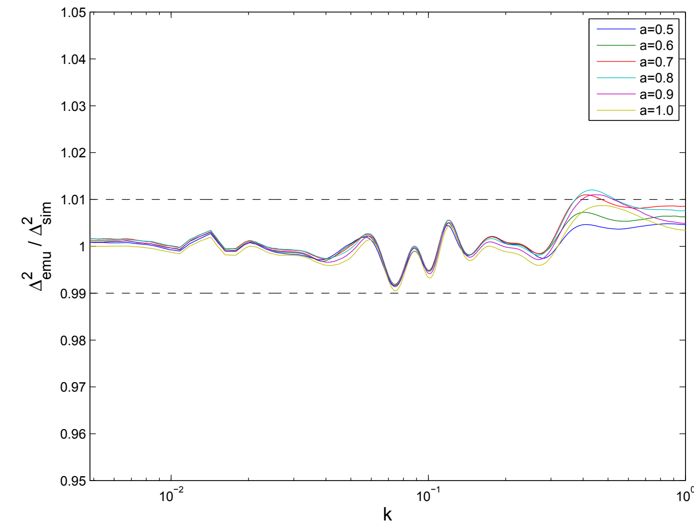

Fig. 2. Ratio of the emulator prediction to the smoothed simulated power spectrum for a ΛCDM cosmology at six values of the scale factor. The error exceeds 1% very slightly in only one part of the domain for the scale factor values, a = 0.7, 0.8, and 0.9.

Future efforts will be targeted to increasing emulation dynamic range, adding more cosmological parameters (dimensions), and constructing emulators for other observables such as the weak lensing shear power spectrum and the cluster mass function. Although our work has shown that emulation is a powerful technique, building successful emulators is technically challenging and requires a concerted effort among different communities.

For more information contact Salman Habib at habib@lanl.gov.

[1] K. Heitmann et al., *Astrophys. J. Lett.* **646**, L1 (2006).
[2] S. Habib et al., *Phys. Rev. D* **76,** 083503 (2007).
[3] K. Heitmann et al., *Astrophys. J.*, arXiv:0812.1052 (to appear).
[4] K. Heitmann et al., *Astrophys. J.* **705,** 156 (2009).
[5] E. Lawrence et al., *Astrophys. J.*, arXiv:0912.4490 (to appear).

Funding Acknowledgments
- LANL Directed Research and Development Program
- NASA Theoretical Astrophysics Program
- DOE High Energy Physics